

## SEQUENCE LISTING

### (1) GENERAL INFORMATION:

- (i) APPLICANT: Yang Wang, Michael W. Spellman
  - (ii) TITLE OF INVENTION: O-Fucosyltransferase
  - (iii) NUMBER OF SEQUENCES: 17
  - (iv) CORRESPONDENCE ADDRESS:
    - (A) ADDRESSEE: Genentech, Inc.
    - (B) STREET: 1 DNA Way
    - (C) CITY: South San Francisco
    - (D) STATE: California
    - (E) COUNTRY: USA
    - (F) ZIP: 94080
  - (v) COMPUTER READABLE FORM:
    - (A) MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
    - (B) COMPUTER: IBM PC compatible
    - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
    - (D) SOFTWARE: WinPatin (Genentech)
  - (vi) CURRENT APPLICATION DATA:
    - (A) APPLICATION NUMBER: Unassigned
    - (B) FILING DATE: 26-Nov-1997
    - (C) CLASSIFICATION:
  - (vii) PRIOR APPLICATION DATA:
    - (A) APPLICATION NUMBER: 08/792498
    - (B) FILING DATE: 31 January 1997
  - (viii) ATTORNEY/AGENT INFORMATION:
    - (A) NAME: Svoboda, Craig G.
    - (B) REGISTRATION NUMBER: 39,044
    - (C) REFERENCE/DOCKET NUMBER: P1041P1
  - (ix) TELECOMMUNICATION INFORMATION:
    - (A) TELEPHONE: 650/225-1489
    - (B) TELEFAX: 650/952-9881
- (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1514 base pairs  
(B) TYPE: Nucleic Acid  
(C) STRANDEDNESS: Single  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGCCCCGCGG GCTCCTGGGA CCCGGCCGGT TACCTGCTCT ACTGCCCCTG 50  
CATGGGGCGC TTTGGGAACC AGGCCGATCA CTTCTTGGGC TCTCTGGCAT 100  
TTGCAAAGCT GCTAAACCGT ACCTTGCTG TCCCTCCTTG GATTGAGTAC 150  
CAGCATCACA AGCCTCCTTT CACCAACCTC CATGTGTCCT ACCAGAAGTA 200  
CTTCAAGCTG GAGCCCCTCC AGGCTTACCA TCGGGTCATC AGCTTGAGAG 250  
ATTTTCATGGA GAAGCTGGCA CCCACCCACT GGCCCCCTGA GAAGCGGGTG 300  
GCATACTGCT TTGAGGTGGC AGCCCAGCGA AGCCCAGATA AGAAGACGTG 350  
CCCCATGAAG GAAGGAAACC CCTTTGGCCC ATTCTGGGAT CAGTTTCATG 400  
TGAGTTTCAA CAAGTCGGAG CTTTTTACAG GCATTTTCCTT CAGTGCTTCC 450  
TACAGAGAAC AATGGAGCCA GAGATTTTCT CCAAAGGAAC ATCCGGTGCT 500  
TGCCCTGCCA GGAGCCCCAG CCCAGTTCCC CGTCCTAGAA GAACACAGGC 550  
CACTACAGAA GTACATGGTA TGGTCAGACG AAATGGTGAA GACGGGAGAG 600  
GCCCAGATTC ATGCCACCT TGTCCGGCCC TATGTGGGCA TTCATCTGCG 650  
CATTGGCTCT GACTGGAAGA ACGCCTGTGC CATGCTGAAG GACGGGACTG 700  
CAGGCTCGCA CTTTCATGGCC TCTCCGCAGT GTGTGGGCTA CAGCCGCAGC 750  
ACAGCGGCCC CCCTCACGAT GACTATGTGC CTGCCTGACC TGAAGGAGAT 800  
CCAGAGGGCT GTGAAGCTCT GGGTGAGGTC GCTGGATGCC CAGTCGGTCT 850

ACGTTGCTAC TGATTCCGAG AGTTATGTGC CTGAGCTCCA ACAGCTCTTC 900  
 AAAGGGAAGG TGAAGGTGGT GAGCCTGAAG CCTGAGGTGG CCCAGGTCGA 950  
 CCTGTACATC CTCGGCCAAG CCGACCACTT TATTGGCAAC TGTGTCTCCT 1000  
 CCTTCACTGC CTTTGTGAAG CGGGAGCGGG ACCTCCAGGG GAGGCCGTCT 1050  
 TCTTTCTTCG GCATGGACAG GCCCCCTAAG CTGCGGGACG AGTTCTGATT 1100  
 CTGGCCGGAG CACCAGACCC TCTGATCCTG GAGGGACCAG AGTCTGAGCT 1150  
 GGTCCCTTCCA GCCAGGCCTG GCAGCCAGAG GTGCTCCGGG ATTGCAAAC 1200  
 CCTCTTCTCA CCTGCCAAAG ATGGAGAAGA GTGCCAGGGA CCCCTCAAGG 1250  
 AGGGAGACGC TCCATATCCC AGGGCATAGG ACTTGCAGGT TCCTAGGAGC 1300  
 AGGAGCATCT CCCATCGCAC GTGCTTTCTG CTCTTCTGGG AATTTCTCAC 1350  
 ACTGGCAAAG CAGTCCAGCC TCCGTCTTCT GGTCCACTCT GCTCTGAGCA 1400  
 GCCTGGGATG CTGAACTCTT CAGAGAGATT TTTTATAGA GAGATTTCTA 1450  
 TAATTTTGAT ACAAGGTCAT GACTATCCTA GAACTCTCTG TGGTTTTTGA 1500  
 AAATCATTGA ATTC 1514

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 365 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Ala	Gly	Ser	Trp	Asp	Pro	Ala	Gly	Tyr	Leu	Leu	Tyr	Cys
1				5					10					15
Pro	Cys	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe	Leu	Gly
				20					25					30

Ser	Leu	Ala	Phe	Ala	Lys	Leu	Leu	Asn	Arg	Thr	Leu	Ala	Val	Pro	35	40	45
Pro	Trp	Ile	Glu	Tyr	Gln	His	His	Lys	Pro	Pro	Phe	Thr	Asn	Leu	50	55	60
His	Val	Ser	Tyr	Gln	Lys	Tyr	Phe	Lys	Leu	Glu	Pro	Leu	Gln	Ala	65	70	75
Tyr	His	Arg	Val	Ile	Ser	Leu	Glu	Asp	Phe	Met	Glu	Lys	Leu	Ala	80	85	90
Pro	Thr	His	Trp	Pro	Pro	Glu	Lys	Arg	Val	Ala	Tyr	Cys	Phe	Glu	95	100	105
Val	Ala	Ala	Gln	Arg	Ser	Pro	Asp	Lys	Lys	Thr	Cys	Pro	Met	Lys	110	115	120
Glu	Gly	Asn	Pro	Phe	Gly	Pro	Phe	Trp	Asp	Gln	Phe	His	Val	Ser	125	130	135
Phe	Asn	Lys	Ser	Glu	Leu	Phe	Thr	Gly	Ile	Ser	Phe	Ser	Ala	Ser	140	145	150
Tyr	Arg	Glu	Gln	Trp	Ser	Gln	Arg	Phe	Ser	Pro	Lys	Glu	His	Pro	155	160	165
Val	Leu	Ala	Leu	Pro	Gly	Ala	Pro	Ala	Gln	Phe	Pro	Val	Leu	Glu	170	175	180
Glu	His	Arg	Pro	Leu	Gln	Lys	Tyr	Met	Val	Trp	Ser	Asp	Glu	Met	185	190	195
Val	Lys	Thr	Gly	Glu	Ala	Gln	Ile	His	Ala	His	Leu	Val	Arg	Pro	200	205	210
Tyr	Val	Gly	Ile	His	Leu	Arg	Ile	Gly	Ser	Asp	Trp	Lys	Asn	Ala	215	220	225
Cys	Ala	Met	Leu	Lys	Asp	Gly	Thr	Ala	Gly	Ser	His	Phe	Met	Ala	230	235	240

Ser	Pro	Gln	Cys	Val	Gly	Tyr	Ser	Arg	Ser	Thr	Ala	Ala	Pro	Leu	
				245					250					255	
Thr	Met	Thr	Met	Cys	Leu	Pro	Asp	Leu	Lys	Glu	Ile	Gln	Arg	Ala	
				260					265					270	
Val	Lys	Leu	Trp	Val	Arg	Ser	Leu	Asp	Ala	Gln	Ser	Val	Tyr	Val	
				275					280					285	
Ala	Thr	Asp	Ser	Glu	Ser	Tyr	Val	Pro	Glu	Leu	Gln	Gln	Leu	Phe	
				290					295					300	
Lys	Gly	Lys	Val	Lys	Val	Val	Ser	Leu	Lys	Pro	Glu	Val	Ala	Gln	
				305					310					315	
Val	Asp	Leu	Tyr	Ile	Leu	Gly	Gln	Ala	Asp	His	Phe	Ile	Gly	Asn	
				320					325					330	
Cys	Val	Ser	Ser	Phe	Thr	Ala	Phe	Val	Lys	Arg	Glu	Arg	Asp	Leu	
				335					340					345	
Gln	Gly	Arg	Pro	Ser	Ser	Phe	Phe	Gly	Met	Asp	Arg	Pro	Pro	Lys	
				350					355					360	
Leu	Arg	Asp	Glu	Phe											
				365											

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 61 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Arg	Leu	Ala	Gly	Ser	Trp	Asp	Leu	Ala	Gly	Tyr	Leu	Leu	Tyr	Xaa	
1				5					10					15	
Pro	Xaa	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe	Leu	Gly	
				20					25					30	



AGCCAGAGAT TTTCTCCAAA GGAACATCCG GTGCTTGCCC TGCCAGGAGC 650  
 CCCAGCCCAG TTCCCCGTCC TAGAGGAACA CAGGCCACTA CAGAAGTACA 700  
 TGGTATGGTC AGACGAAATG GTGAAGACGG GAGAGGCCCA GATTCATGCC 750  
 CACCTTGTCC GGCCCTATGT GGGCATTTCAT CTGCGCATTG GCTCTGACTG 800  
 GAAGAACGCC TGTGCCATGC TGAAGGACGG GACTGCAGGC TCGCACTTCA 850  
 TGGCCTCTCC GCAGTGTGTG GGCTACAGCC GCAGCACAGC GGCCCCCCTC 900  
 ACGATGACTA TGTGCCTGCC TGACCTGAAG GAGATCCAGA GGGCTGTGAA 950  
 GCTCTGGGTG AGGTCGCTGG ATGCCCAGTC GGTCTACGTT GCTACTGATT 1000  
 CCGAGAGTTA TGTGCCTGAG CTCCAACAGC TCTTCAAAGG GAAGGTGAAG 1050  
 GTGGTGAGCC TGAAGCCTGA GGTGGCCCAG GTCGACCTGT ACATCCTCGG 1100  
 CCAAGCCGAC CACTTTATTG GCAACTGTGT CTCCTCCTTC ACTGCCTTTG 1150  
 TGAAGCGGGA GCGGGACCTC CAGGGGAGGC CGTCTTCTTT CTTGCGCATG 1200  
 GACAGGCCCC CTAAGCTGCG GGACGAGTTC TGATTCTGGC CGGAGCACCA 1250  
 GACCCTCTGA TCCTGGAGGG ACCAGAGTCT GAGCTGGTCC TTCCAGCCAG 1300

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11284 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAGCTTTACT CGTAAAGCGA GTTGAAGGAT CATATTTAGT TGCGTTTATG 50  
 AGATAAGATT GAAAGCACGT GTAAAATGTT TCCCGCGCGT TGGCACAACT 100

ATTTACAATG CGGCCAAGTT ATAAAAGATT CTAATCTGAT ATGTTTTTAAA 150  
 ACACCTTTGC GGGCCGAGTT GTTTGCGTAC GTGACTAGCG AAGAAGATGT 200  
 GTGGACCGCA GAACAGATAG TAAAACAAAA CCCTAGTATT GGAGCAATAA 250  
 TCGATTTAAC CAACACGTCT AAATATTATG ATGGTGTGCA TTTTTTGCGG 300  
 GCGGGCCTGT TATACAAAAA AATTCAAGTA CCTGGCCAGA CTTTGCCGCC 350  
 TGAAAGCATA GTTCAAGAAT TTATTGACAC GGTAAAAGAA TTTACAGAAA 400  
 AGTGTCCCGG CATGTTGGTG GCGGTGCACT GCACACACGG TATTAATCGC 450  
 ACCGGTTACA TGGTGTGCAG ATATTTAATG CACACCCTGG GTATTGCGCC 500  
 GCAGGAAGCC ATAGATAGAT TCGAAAAAGC CAGAGGTCAC AAAATTGAAA 550  
 GACAAAATTA CGTTCAAGAT TTATTAATTT AATTAATATT ATTTGCATTC 600  
 TTTAACAAAT ACTTTATCCT ATTTTCAAAT TGTTGCGCTT CTTCCAGCGA 650  
 ACCAAACTA TGCTTCGCTT GCTCCGTTTA GCTTGTAGCC GATCAGTGGC 700  
 GTTGTTCCAA TCGACGGTAG GATTAGGCCG GATATTCTCC ACCACAATGT 750  
 TGGCAACGTT GATGTTACGT TTATGCTTTT GGTTTTCCAC GTACGTCTTT 800  
 TGGCCGGTAA TAGCCGTAAA CGTAGTGCCG TCGCGCGTCA CGCACAACAC 850  
 CGGATGTTTG CGCTTGTCCG CGGGGTATTG AACCGCGCGA TCCGACAAAT 900  
 CCACCACTTT GGCAACTAAA TCGGTGACCT GCGCGTCTTT TTTCTGCATT 950  
 ATTTCGTCTT TCTTTTGCAT GGTTTCCTGG AAGCCGGTGT ACATGCGGTT 1000  
 TAGATCAGTC ATGACGCGCG TGACCTGCAA ATCTTTGGCC TCGATCTGCT 1050  
 TGTCTTGAT GGCAACGATG CGTTCAATAA ACTCTTGTTT TTTAACAAAGT 1100  
 TCCTCGGTTT TTTGCGCCAC CACCGCTTGC AGCGCGTTTG TGTGCTCGGT 1150

GAATGTCGCA ATCAGCTTAG TCACCAACTG TTTGCTCTCC TCCTCCCGTT 1200  
GTTTGATCGC GGGATCGTAC TTGCCGGTGC AGAGCACTTG AGGAATTACT 1250  
TCTTCTAAAA GCCATTCTTG TAATTCTATG GCGTAAGGCA ATTTGGACTT 1300  
CATAATCAGC TGAATCACGC CGGATTTAGT AATGAGCACT GTATGCGGCT 1350  
GCAAATACAG CGGGTCGCCC CTTTTCACGA CGCTGTTAGA GGTAGGGCCC 1400  
CCATTTTGGG TGGTCTGCTC AAATAACGAT TTGTATTTAT TGTCTACATG 1450  
AACACGTATA GCTTTATCAC AAACGTGATA TTTTAACTG TTAGCGACGT 1500  
CCTTGGCCAC GAACCGGACC TGTGCGTGC GCTCTAGCAC GTACCGCAGG 1550  
TTGAACGTAT CTTCTCCAAA TTTAAATTCT CCAATTTTAA CGCGAGCCAT 1600  
TTTGATACAC GTGTGTCGAT TTTGCAACAA CTATTGTTTT TTAACGCAAA 1650  
CTAAACTTAT TGTGGTAAGC AATAATTAAA TATGGGGGAA CATGCGCCGC 1700  
TACAACACTC GTCGTTATGA ACGCAGACGG CGCCGGTCTC GGC GCAAGCG 1750  
GCTAAAACGT GTTGCGCGTT CAACGCGGCA AACATCGCAA AAGCCAATAG 1800  
TACAGTTTTG ATTTGCATAT TAACGGCGAT TTTTAAATT ATCTTATTTA 1850  
ATAAATAGTT ATGACGCCTA CAACTCCCCG CCCGCGTTGA CTCGCTGCAC 1900  
CTCGAGCAGT TCGTTGACGC CTTCTCCGT GTGGCCGAAC ACGTCGAGCG 1950  
GGTGGTCGAT GACCAGCGGC GTGCCGCACG CGACGCACAA GTATCTGTAC 2000  
ACCGAATGAT CGTCGGGCGA AGGCACGTCG GCCTCCAAGT GGCAATATTG 2050  
GCAAATTCGA AAATATATAC AGTTGGGTTG TTTGCGCATA TCTATCGTGG 2100  
CGTTGGGCAT GTACGTCCGA ACGTTGATTT GCATGCAAGC CGAAATTAAA 2150  
TCATTGCGAT TAGTGCGATT AAAACGTTGT ACATCCTCGC TTTTAATCAT 2200  
GCCGTCGATT AAATCGCGCA ATCGAGTCAA GTGATCAAAG TGTGGAATAA 2250

TGTTTTCTTT GTATTCCCGA GTCAAGCGCA GCGCGTATTT TAACAAACTA 2300  
 GCCATCTTGT AAGTTAGTTT CATTTAATGC AACTTTATCC AATAATATAT 2350  
 TATGTATCGC ACGTCAAGAA TTAACAATGC GCCCGTTGTC GCATCTCAAC 2400  
 ACGACTATGA TAGAGATCAA ATAAAGCGCG AATTAAATAG CTTGCGACGC 2450  
 AACGTGCACG ATCTGTGCAC GCGTTCGGC ACGAGCTTTG ATTGTAATAA 2500  
 GTTTTTACGA AGCGATGACA TGACCCCGT AGTGACAACG ATCACGCCCA 2550  
 AAAGAACTGC CGACTACAAA ATTACCGAGT ATGTCGGTGA CGTTAAAACT 2600  
 ATTAAGCCAT CCAATCGACC GTTAGTCGAA TCAGGACCGC TGGTGCGAGA 2650  
 AGCCGCGAAG TATGGCGAAT GCATCGTATA ACGTGTGGAG TCCGCTCATT 2700  
 AGAGCGTCAT GTTTAGACAA GAAAGCTACA TATTTAATTG ATCCCGATGA 2750  
 TTTTATTGAT AAATTGACCC TAACTCCATA CACGGTATTC TACAATGGCG 2800  
 GGGTTTTGGT CAAAATTTCC GGACTGCGAT TGTACATGCT GTTAACGGCT 2850  
 CCGCCCACTA TTAATGAAAT TAAAAATTCC AATTTTAAAA AACGCAGCAA 2900  
 GAGAAACATT TGTATGAAAG AATGCGTAGA AGGAAAGAAA AATGTCGTCG 2950  
 ACATGCTGAA CAACAAGATT AATATGCCTC CGTGTATAAA AAAAATATTG 3000  
 AACGATTTGA AAGAAAACAA TGTACCGCGC GCGGTATGT ACAGGAAGAG 3050  
 GTTTATACTA AACTGTTACA TTGCAAACGT GGTTCGTGT GCCAAGTGTG 3100  
 AAAACCGATG TTTAATCAAG GCTCTGACGC ATTTCTACAA CCACGACTCC 3150  
 AAGTGTGTGG GTGAAGTCAT GCATCTTTTA ATCAAATCCC AAGATGTGTA 3200  
 TAAACCACCA AACTGCCAAA AAATGAAAAC TGTCGACAAG CTCTGTCCGT 3250  
 TTGCTGGCAA CTGCAAGGGT CTCAATCCTA TTTGTAATTA TTGAATAATA 3300

AAACAATTAT AAATGCTAAA TTTGTTTTTTT ATTAACGATA CAAACCAAAC 3350  
 GCAACAAGAA CATTTGTAGT ATTATCTATA ATTGAAAACG CGTAGTTATA 3400  
 ATCGCTGAGG TAATATTTAA AATCATTTTC AAATGATTCA CAGTTAATTT 3450  
 GCGACAATAT AATTTTATTT TCACATAAAC TAGACGCCTT GTCGTCTTCT 3500  
 TCTTCGTATT CCTTCTCTTT TTCATTTTTC TCCTCATAAA AATTAACATA 3550  
 GTTATTATCG TATCCATATA TGTATCTATC GTATAGAGTA AATTTTTTGT 3600  
 TGTCAATAAT ATATATGTCT TTTTAAATGG GGTGTATAGT ACCGCTGCGC 3650  
 ATAGTTTTTC TGTAATTTAC AACAGTGCTA TTTTCTGGTA GTTCTTCGGA 3700  
 GTGTGTTGCT TTAATTATTA AATTTATATA ATCAATGAAT TTGGGATCGT 3750  
 CGGTTTTGTA CAATATGTTG CCGGCATAGT ACGCAGCTTC TTCTAGTTCA 3800  
 ATTACACCAT TTTT TAGCAG CACCGGATTA ACATAACTTT CCAAATGTT 3850  
 GTACGAACCG TTAAACAAAA ACAGTTCACC TCCCTTTTCT ATACTATTGT 3900  
 CTGCGAGCAG TTGTTTGTTG TTAAAAATAA CAGCCATTGT AATGAGACGC 3950  
 ACAAACTAAT ATCACAACT GGAAATGTCT ATCAATATAT AGTTGCTGAT 4000  
 ATCATGGAGA TAATTAAAT GATAACCATC TCGCAAATAA ATAAGTATTT 4050  
 TACTGTTTTTC GTAACAGTTT TGTAATAAAA AAACCTATAA ATATTCCGGA 4100  
 TTATTCATAC CGTCCCACCA TCGGGCGCGG ATCAGATCCA TGGCCAAGTT 4150  
 CCTGGTCAAC GTGGCCCTGC TGCTGCTGCT GCTGCTGCTG TCCGGAGCCT 4200  
 GGGCCCATAT GAGATCCCAT CACCATCACC ATCACATGCC CGCGGGCTCC 4250  
 TGGGACCCGG CCGGTTACCT GCTCTACTGC CCCTGCATGG GGCGCTTTGG 4300  
 GAACCAGGCC GATCACTTCT TGGGCTCTCT GGCATTTGCA AAGCTGCTAA 4350  
 ACCGTACCTT GGCTGTCCCT CTTGGATTG AGTACCAGCA TCACAAGCCT 4400

CCTTTCACCA ACCTCCATGT GTCCTACCAG AAGTACTTCA AGCTGGAGCC 4450  
CCTCCAGGCT TACCATCGGG TCATCAGCTT GGAGGATTTC ATGGAGAAGC 4500  
TGGCACCCAC CCACTGGCCC CCTGAGAAGC GGGTGGCATA CTGCTTTGAG 4550  
GTGGCAGCCC AGCGAAGCCC AGATAAGAAG ACGTGCCCCA TGAAGGAAGG 4600  
AAACCCCTTT GGCCCATTTCT GGGATCAGTT TCATGTGAGT TTCAACAAGT 4650  
CGGAGCTTTT TACAGGCATT TCCTTCAGTG CTTCTACAG AGAACAATGG 4700  
AGCCAGAGAT TTTCTCCAAA GGAACATCCG GTGCTTGCCC TGCCAGGAGC 4750  
CCCAGCCCAG TTCCCCGTCC TAGAGGAACA CAGGCCACTA CAGAAGTACA 4800  
TGGTATGGTC AGACGAAATG GTGAAGACGG GAGAGGCCCA GATTCATGCC 4850  
CACCTTGTCC GGCCCTATGT GGGCATT CAT CTGCGCATTG GCTCTGACTG 4900  
GAAGAACGCC TGTGCCATGC TGAAGGACGG GACTGCAGGC TCGCACTTCA 4950  
TGGCCTCTCC GCAGTGTGTG GGCTACAGCC GCAGCACAGC GGCCCCCCTC 5000  
ACGATGACTA TGTGCCTGCC TGACCTGAAG GAGATCCAGA GGGCTGTGAA 5050  
GCTCTGGGTG AGGTCGCTGG ATGCCCAGTC GGTCTACGTT GCTACTGATT 5100  
CCGAGAGTTA TGTGCCTGAG CTCCAACAGC TCTTCAAAGG GAAGGTGAAG 5150  
GTGGTGAGCC TGAAGCCTGA GGTGGCCCAG GTCGACCTGT ACATCCTCGG 5200  
CCAAGCCGAC CACTTTATTG GCAACTGTGT CTCCTCCTTC ACTGCCTTTG 5250  
TGAAGCGGGA GCGGGACCTC CAGGGGAGGC CGTCTTCTTT CTTCGGCATG 5300  
GACAGGCCCC CTAAGCTGCG GGACGAGTTC TGATTCTGGC CGGAGCACCA 5350  
GACCTCTGA TCCTGGAGGG ACCAGAGTCT GAGCTGGTCC TTCCAGCCAG 5400  
GCCTGGCAGC CAGAGGTGCT CCGGGATTGC AAACCTCCTCT TCTCACCTGC 5450

CAAAGATGGA GAAGAGTGCC AGGGACCCCT CAAGGAGGGA GACGCTCCAT 5500  
 ATCCCAGGGC ATAGGACTTG CAGGTTCCCTA GGAGCAGGAG CATCTCCCAT 5550  
 CGCACGTGCT TTCTGCTCTT CTGGGAATTT CTCACACTGG CAAAGCAGTC 5600  
 CAGCCTCCGT CTTCTGGTCC ACTCTGCTCT GAGCAGCCTG GGATGCTGAA 5650  
 CTCTTCAGAG AGATTTTTTT ATAGAGAGAT TTCTATAATT TTGATACAAG 5700  
 GTCATGACTA TCCTAGAACT CTCTGTGGTT TTTGAAAATC ATTGAATTCC 5750  
 TGCAGCCCGG GGGATCCACT AGTTCTAGTT CTAGAGCGGC CGCTCCAGAA 5800  
 TTCTAGAAGG TACCCGGGAT CCTTTCCTGG GACCCGGCAA GAACCAAAAA 5850  
 CTCACTCTCT TCAAGGAAAT CCGTAATGTT AAACCCGACA CGATGAAGCT 5900  
 TGTCGTTGGA TGGAAAGGAA AAGAGTTCTA CAGGGAAACT TGGACCCGCT 5950  
 TCATGGAAGA CAGCTTCCCC ATTGTTAACG ACCAAGAAGT GATGGATGTT 6000  
 TTCCTTGTTG TCAACATGCG TCCCACTAGA CCAACCGTT GTTACAAATT 6050  
 CCTGGCCCAA CACGCTCTGC GTTGCGACCC CGACTATGTA CCTCATGACG 6100  
 TGATTAGGAT CGTCGAGCCT TCATGGGTGG GCAGCAACAA CGAGTACCGC 6150  
 ATCAGCCTGG CTAAGAAGGG CGGCGGCTGC CCAATAATGA ACCTTCACTC 6200  
 TGAGTACACC AACTCGTTTCG AACAGTTCAT CGATCGTGTC ATCTGGGAGA 6250  
 ACTTCTACAA GCCCATCGTT TACATCGGTA CCGACTCTGC TGAAGAGGAG 6300  
 GAAATTCTCC TTGAAGTTTC CCTGGTGTTT AAAGTAAAGG AGTTTGCACC 6350  
 AGACGCACCT CTGTTCACTG GTCCGGCGTA TTAAAACACG ATACATTGTT 6400  
 ATTAGTACAT TTATTAAGCG CTAGATTCTG TGC GTTGTGTTG ATTTACAGAC 6450  
 AATTGTTGTA CGTATTTTAA TAATTCATTA AATTTATAAT CTTTAGGGTG 6500  
 GTATGTTAGA GCGAAAATCA AATGATTTTC AGCGTCTTTA TATCTGAATT 6550

TAAATATTAA ATCCTCAATA GATTTGTAAA ATAGGTTTCG ATTAGTTTCA 6600  
AACAGGGTT GTTTTTCCGA ACCGATGGCT GGACTATCTA ATGGATTTTC 6650  
GCTCAACGCC ACAAACTTG CCAAATCTTG TAGCAGCAAT CTAGCTTTGT 6700  
CGATATTCGT TTGTGTTTTG TTTTGTAAATA AAGGTTTCGAC GTCGTTCAAA 6750  
ATATTATGCG CTTTTGTATT TCTTTCATCA CTGTCGTTAG TGTACAATTG 6800  
ACTCGACGTA AACACGTAA ATAAAGCTTG GACATATTTA ACATCGGGCG 6850  
TGTTAGCTTT ATTAGGCCGA TTATCGTCGT CGTCCCAACC CTCGTCGTTA 6900  
GAAGTTGCTT CCGAAGACGA TTTTGCCATA GCCACACGAC GCCTATTAAT 6950  
TGTGTCGGCT AACACGTCCG CGATCAAATT TGTTAGTTGAG CTTTTTGGAA 7000  
TTATTTCTGA TTGCGGGCGT TTTTGGGCGG GTTTCATCT AACTGTGCCC 7050  
GATTTTAATT CAGACAACAC GTTAGAAAGC GATGGTGCAG GCGGTGGTAA 7100  
CATTTAGAC GGCAAATCTA CTAATGGCGG CCGTGGTGGA GCTGATGATA 7150  
AATCTACCAT CGGTGGAGGC GCAGGCGGGG CTGGCGGCGG AGGCGGAGGC 7200  
GGAGGTGGTG GCGGTGATGC AGACGGCGGT TTAGGCTCAA ATGTCTCTTT 7250  
AGGCAACACA GTCGGCACCT CAACTATTGT ACTGGTTTCG GCGCCGTTT 7300  
TTGGTTTGAC CGGTCTGAGA CGAGTGCGAT TTTTTCGTT TCTAATAGCT 7350  
TCCAACAATT GTTGTCTGTC GTCTAAAGGT GCAGCGGGTT GAGGTTCCGT 7400  
CGGCATTGGT GGAGCGGGCG GCAATTCAGA CATCGATGGT GGTGGTGGTG 7450  
GTGGAGGCGC TGGAATGTTA GGCACGGGAG AAGGTGGTGG CGGCGGTGCC 7500  
GCCGGTATAA TTTGTTCTGG TTTAGTTTGT TCGCGCACGA TTGTGGGCAC 7550  
CGGCGCAGGC GCCGCTGGCT GCACAACGGA AGGTCGTCTG CTTGAGGCA 7600

GCGCTTGGGG TGGTGGCAAT TCAATATTAT AATTGGAATA CAAATCGTAA 7650  
 AAATCTGCTA TAAGCATTGT AATTTGCTA TCGTTTACCG TGCCGATATT 7700  
 TAACAACCGC TCAATGTAAG CAATTGTATT GTAAAGAGAT TGTCTCAAGC 7750  
 TCCGCACGCC GATAACAAGC CTTTTCATTT TTACTACAGC ATTGTAGTGG 7800  
 CGAGACACTT CGCTGTCGTC GACGTACATG TATGCTTTGT TGTCAAAAAC 7850  
 GTCGTTGGCA AGCTTTAAAA TATTTAAAAG AACATCTCTG TTCAGCACCA 7900  
 CTGTGTTGTC GTAAATGTTG TTTTGTATAA TTTGCGCTTC CGCAGTATCG 7950  
 ACACGTTCAA AAAATTGATG CGCATCAATT TTGTTGTTCC TATTATTGAA 8000  
 TAAATAAGAT TGTACAGATT CATATCTACG ATTCGTCATG GCCACCACAA 8050  
 ATGCTACGCT GCAAACGCTG GTACAATTTT ACGAAAACGTC CAAAACGTC 8100  
 AAAACTCGGT ATAAAATAAT CAACGGGCGC TTTGGCAAAA TATCTATTTT 8150  
 ATCGCACAAG CCCACTAGCA AATTGTATTT GCAGAAAACA ATTTGCGCGC 8200  
 ACAATTTTAA CGCTGACGAA ATAAAAGTTC ACCAGTTAAT GAGCGACCAC 8250  
 CCAAATTTTA TAAAAATCTA TTTTAATCAC GGTTCATCA ACAACCAAGT 8300  
 GATCGTGATG GACTACATTG ACTGTCCCGA TTTATTTGAA AACTACAAA 8350  
 TTAAAGGCGA GCTTTCGTAC CAACTTGTTA GCAATATTAT TAGACAGCTG 8400  
 TGTGAAGCGC TCAACGATTT GCACAAGCAC AATTCATAC ACAACGACAT 8450  
 AAAACTCGAA AATGTCTTAT ATTTGGAAGC ACTTGATCGC GTGTATGTTT 8500  
 GCGATTACGG ATTGTGCAAA CACGAAAAC TACTTAGCGT GCACGACGGC 8550  
 ACGTTGGAGT ATTTTAGTCC GGAAAAAATT CGACACACAA CTATGCACGT 8600  
 TTCGTTTGAC TGGTACGCGG CGTGTTAACA TACAAGTTGC TAACCGGCGG 8650  
 TTCGTAATCA TGGTCATAGC TGTTCCTGT GTGAAATTGT TATCCGCTCA 8700

CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTA AGCCTGGGGT 8750  
 GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCCG 8800  
 TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC 8850  
 GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC 8900  
 ACTGACTCGC TGCCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA 8950  
 CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA 9000  
 AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC 9050  
 GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG AGCATCACAA 9100  
 AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCGACAGGA CTATAAAGAT 9150  
 ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCCTCTCC TGTTCCGACC 9200  
 CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC 9250  
 GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTC 9300  
 GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC 9350  
 GCCTTATCCG GTAACATCG TCTTGAGTCC AACCCGGTAA GACACGACTT 9400  
 ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG 9450  
 TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT 9500  
 AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG 9550  
 AAAAAGAGTT GGTAAGTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG 9600  
 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT 9650  
 CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA 9700  
 AACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA 9750

CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA 9800  
TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC 9850  
TATCTCAGCG ATCTGTCTAT TTCGTTTCATC CATAGTTGCC TGA CTCCCCG 9900  
TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT 9950  
GCAATGATAC CGCGAGACCC ACGCTCACC GCTCCAGATT TATCAGCAAT 10000  
AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT 10050  
CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG AGTAAGTAGT 10100  
TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT 10150  
GGTGTACGCG TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC 10200  
GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAAA AGCGGTTAGC 10250  
TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC 10300  
ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG 10350  
TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA 10400  
TAGTGTATGC GCGGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA 10450  
TACCGCGCCA CATAGCAGAA CTTTAAAAGT GTCATCATT GGAAAACGTT 10500  
CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG 10550  
ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TACTTTTAC 10600  
CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAATGCC GCAAAAAGG 10650  
GAATAAGGGC GACACGGAAA TGTGAATAC TCATACTCTT CCTTTTCAA 10700  
TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT 10750  
TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCGCGC ACATTTCCCC 10800  
GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC 10850

TATAAAAATA GCGTATCAC GAGGCCCTTT CGTCTCGCGC GTTTCGGTGA 10900  
 TGACGGTGAA AACCTCTGAC ACATGCAGCT CCCGGAGACG GTCACAGCTT 10950  
 GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG CGCGTCAGCG 11000  
 GGTGTTGGCG GGTGTCGGGG CTGGCTTAAC TATGCGGCAT CAGAGCAGAT 11050  
 TGTACTGAGA GTGCACCATA TATGCGGTGT GAAATACCGC ACAGATGCGT 11100  
 AAGGAGAAAA TACCGCATCA GGCGCCATTC GCCATTCAGG CTGCGCAACT 11150  
 GTTGGGAAGG GCGATCGGTG CGGGCCTCTT CGCTATTACG CCAGCTGGCG 11200  
 AAAGGGGGAT GTGCTGCAAG GCGATTAAGT TGGGTAACGC CAGGGTTTTTC 11250  
 CCAGTCACGA CGTTGTAAAA CGACGGCCAG TGCC 11284

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 397 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met	Ala	Lys	Phe	Leu	Val	Asn	Val	Ala	Leu	Leu	Leu	Leu	Leu	Leu	1	5	10	15
Leu	Leu	Ser	Gly	Ala	Trp	Ala	His	Met	Arg	Ser	His	His	His	His	20	25	30	
His	His	Met	Pro	Ala	Gly	Ser	Trp	Asp	Pro	Ala	Gly	Tyr	Leu	Leu	35	40	45	
Tyr	Cys	Pro	Cys	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe	50	55	60	
Leu	Gly	Ser	Leu	Ala	Phe	Ala	Lys	Leu	Leu	Asn	Arg	Thr	Leu	Ala	65	70	75	

Val	Pro	Pro	Trp	Ile	Glu	Tyr	Gln	His	His	Lys	Pro	Pro	Phe	Thr	80	85	90
Asn	Leu	His	Val	Ser	Tyr	Gln	Lys	Tyr	Phe	Lys	Leu	Glu	Pro	Leu	95	100	105
Gln	Ala	Tyr	His	Arg	Val	Ile	Ser	Leu	Glu	Asp	Phe	Met	Glu	Lys	110	115	120
Leu	Ala	Pro	Thr	His	Trp	Pro	Pro	Glu	Lys	Arg	Val	Ala	Tyr	Cys	125	130	135
Phe	Glu	Val	Ala	Ala	Gln	Arg	Ser	Pro	Asp	Lys	Lys	Thr	Cys	Pro	140	145	150
Met	Lys	Glu	Gly	Asn	Pro	Phe	Gly	Pro	Phe	Trp	Asp	Gln	Phe	His	155	160	165
Val	Ser	Phe	Asn	Lys	Ser	Glu	Leu	Phe	Thr	Gly	Ile	Ser	Phe	Ser	170	175	180
Ala	Ser	Tyr	Arg	Glu	Gln	Trp	Ser	Gln	Arg	Phe	Ser	Pro	Lys	Glu	185	190	195
His	Pro	Val	Leu	Ala	Leu	Pro	Gly	Ala	Pro	Ala	Gln	Phe	Pro	Val	200	205	210
Leu	Glu	Glu	His	Arg	Pro	Leu	Gln	Lys	Tyr	Met	Val	Trp	Ser	Asp	215	220	225
Glu	Met	Val	Lys	Thr	Gly	Glu	Ala	Gln	Ile	His	Ala	His	Leu	Val	230	235	240
Arg	Pro	Tyr	Val	Gly	Ile	His	Leu	Arg	Ile	Gly	Ser	Asp	Trp	Lys	245	250	255
Asn	Ala	Cys	Ala	Met	Leu	Lys	Asp	Gly	Thr	Ala	Gly	Ser	His	Phe	260	265	270
Met	Ala	Ser	Pro	Gln	Cys	Val	Gly	Tyr	Ser	Arg	Ser	Thr	Ala	Ala	275	280	285

Pro	Leu	Thr	Met	Thr	Met	Cys	Leu	Pro	Asp	Leu	Lys	Glu	Ile	Gln	290	295	300
Arg	Ala	Val	Lys	Leu	Trp	Val	Arg	Ser	Leu	Asp	Ala	Gln	Ser	Val	305	310	315
Tyr	Val	Ala	Thr	Asp	Ser	Glu	Ser	Tyr	Val	Pro	Glu	Leu	Gln	Gln	320	325	330
Leu	Phe	Lys	Gly	Lys	Val	Lys	Val	Val	Ser	Leu	Lys	Pro	Glu	Val	335	340	345
Ala	Gln	Val	Asp	Leu	Tyr	Ile	Leu	Gly	Gln	Ala	Asp	His	Phe	Ile	350	355	360
Gly	Asn	Cys	Val	Ser	Ser	Phe	Thr	Ala	Phe	Val	Lys	Arg	Glu	Arg	365	370	375
Asp	Leu	Gln	Gly	Arg	Pro	Ser	Ser	Phe	Phe	Gly	Met	Asp	Arg	Pro	380	385	390
Pro	Lys	Leu	Arg	Asp	Glu	Phe									395	397	

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5009 base pairs
  - (B) TYPE: Nucleic Acid
  - (C) STRANDEDNESS: Single
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

GAACCAGGCC GATCACTTCT TGGGCTCTCT GGCATTTGCA AAGCTGCTAA 50
ACCGTACCTT GGCTGTCCCT CTTGGATTG AGTACCAGCA TCACAAGCCT 100
CCTTTCACCA ACCTCCATGT GTCCTACCAG AAGTACTTCA AGCTGGAGCC 150
CCTCCAGGCT TACCATCGGG TCATCAGCTT GGAGGATTTC ATGGAGAAGC 200

```

TGGCACCCAC CCACTGGCCC CCTGAGAAGC GGGTGGCATA CTGCTTTGAG 250  
 GTGGCAGCCC AGCGAAGCCC AGATAAGAAG ACGTGCCCCA TGAAGGAAGG 300  
 AAACCCCTTT GGCCCATTTCT GGGATCAGTT TCATGTGAGT TTCAACAAGT 350  
 CGGAGCTTTT TACAGGCATT TCCTTCAGTG CTTCTACAG AGAACAAATGG 400  
 AGCCAGAGAT TTTCTCCAAA GGAACATCCG GTGCTTGCCC TGCCAGGAGC 450  
 CCCAGCCCAG TTCCCCGTCC TAGAGGAACA CAGGCCACTA CAGAAGTACA 500  
 TGGTATGGTC AGACGAAATG GTGAAGACGG GAGAGGCCCA GATTCATGCC 550  
 CACCTTGTCG GGCCCTATGT GGGCATTCAT CTGCGCATTG GCTCTGACTG 600  
 GAAGAACGCC TGTGCCATGC TGAAGGACGG GACTGCAGGC TCGCACTTCA 650  
 TGGCCTCTCC GCAGTGTGTG GGCTACAGCC GCAGCACAGC GGCCCCCCTC 700  
 ACGATGACTA TGTGCCTGCC TGACCTGAAG GAGATCCAGA GGGCTGTGAA 750  
 GCTCTGGGTG AGGTCGCTGG ATGCCCAGTC GGTCTACGTT GCTACTGATT 800  
 CCGAGAGTTA TGTGCCTGAG CTCCAACAGC TCTTCAAAGG GAAGGTGAAG 850  
 GTGGTGAGCC TGAAGCCTGA GGTGGCCCAG GTCGACCTGT ACATCCTCGG 900  
 CCAAGCCGAC CACTTTATTG GCAACTGTGT CTCCTCCTTC ACTGCCTTTG 950  
 TGAAGCGGGA GCGGGACCTC CAGGGGAGGC CGTCTTCTTT CTTCCGCATG 1000  
 GACAGGCCCC CTAAGCTGCG GGACGAGTTC TGATTCTGGC CGGAGCACCA 1050  
 GACCCTCTGA TCCTGGAGGG ACCAGAGTCT GAGCTGGTCC TTCCAGCCAG 1100  
 GCCTGGCAGC CAGAGGTGCT CCGGGATTGC AAATCCTCT TCTCACCTGC 1150  
 CAAAGATGGA GAAGAGTGCC AGGGACCCCT CAAGGAGGGA GACGCTCCAT 1200  
 ATCCCAGGGC ATAGGACTTG CAGGTTCTTA GGAGCAGGAG CATCTCCCAT 1250  
 CGCACGTGCT TTCTGCTCTT CTGGGAATTT CTCACACTGG CAAAGCAGTC 1300

CAGCCTCCGT CTTCTGGTCC ACTCTGCTCT GAGCAGCCTG GGATGCTGAA 1350  
CTCTTCAGAG AGATTTTTTTT ATAGAGAGAT TTCTATAATT TTGATAACAAG 1400  
GTCATGACTA TCCTAGAACT CTCTGTGGTT TTTGAAAATC ATTGAATTCT 1450  
ATTAATGTAG GTACCTAAAG TGACCTTAAC TGAATGTGGA TGAGGCTGGG 1500  
GCTGGTGTGG GTCTTTTGGC TGCTTTTCAA GGTGTCCCCC AATGTGGCCC 1550  
TCAAGAGCCA TCCCCACTGC CTGGCCAGAG CCATTGTTGT CCCCTACTTC 1600  
CTAGGCCATT TCTGGGGCTT GGGGGATGAA TGCTGTCCTG TGCTGTAAAC 1650  
ACTATGCAAA TGGAAGTTAT CGGTTGTGGT GCTGTGCAGC GCTCTGTGGG 1700  
CGACTAAGTG CCACTCACGC AGCATGTTCC TGGCAAGGAG CACATACCAT 1750  
CAAGCCACAC TATCATGGTA TTGTTCTCAC AGTCTTTTGG TGGTTGATGG 1800  
CCACTGCAAA CCTGGCACCA TCAGATCTCT TCTGATCTCT TGCCCCAGTG 1850  
GGGCCTGGTT GGTAAGTGT TGGCATTCGG TTGATATCCA AAGCCTGTTC 1900  
TCCCAGCCGT CCTCCTGCAG CTGGAGCCTT CAGGCCGTAT TCTCACGAGG 1950  
GAACGTTTGC CAAGGCTCTG ACCTCACAGA AGATGCCAG GGCCCAGAAG 2000  
CCATCAGAAT TATCAGTGA GAAGCACCTT TTGACTCTTC CCTTCCAATG 2050  
TAATCTCTGC CAACACCATG AGGCTTAAGG TGCTCTAAGT CATGAGTGTT 2100  
TTGGTCTCAA ATGCTGCAGT TTTAATAATC TGTGACTCCT GAGAGCCCAT 2150  
GGTTTTTTGA CCTTGTGGTT CTAAAATTCC TTGTCTGACC CCTGTAGATC 2200  
TTTTCTTGC CATGTCACCT CCCTTGGCCT TTGATCCTGG AAAGGTGGCA 2250  
GAGCCTCCAC TGAGCCAGGC CCAGAGCTCC TTGCAGTGCC TTCTTCCTTG 2300  
TTTACCTGTG GGAGGAAACA CTTTTTTTGT CAGGGGCAGC CTGGTTCAGA 2350

GCTCAGAGGT CACACTGTAT CAAAGATCTC AAACAGCAAA GTCAGCATTT 2400  
GCTGTATAGA GCTGCCACCC AACTCTAAGC AGGAGAAACT GTACAGAAAG 2450  
GGCTTTGCTA TTTTTCCTT TTGGGAAAAC AATGAAGTGT TTTAAGTCCT 2500  
GGGTGGACTG AGAGATGGTT TGCCTGTCCA GACTTGCTCT CAAGCCTCAT 2550  
CCAGAGAAGG AGCTGCAGAT GAGGGAGCCC GTACACTCCC TGCCACCACT 2600  
AGGTTGTAAG CCTGTAGCTG GCTGGCTGAT TTCATTTTGG AATTCATTTG 2650  
CCATCCACAG CCTTACACTA GGCACACACT TTAGAGTCTG GGGCTCCAGT 2700  
GGGGCCCGCC TAATTTTTTTT TCCCCCAAG ACAGGGCCTT GCTCTGTCTC 2750  
CCAGGCTGGA GTGCAGTGGC ATGATCATGG CTTACTGCAG CCTTGATCTC 2800  
CCAGGCTCAA GCGATCCTTC TGCCTCAGCC TCTCTGGTAG CTGAGACTGC 2850  
ATGCCCAGCT CCAAATCACC TTGATTCATA TCAGCAGTAA TAATCACTTG 2900  
TGTTCTGAAA GAAAGGGCAC CAGAAGTTCT AGCAAAATTC AGTTGTGTTC 2950  
TGTGAGCTAG CACTTTTTTC TCTGACCCAA TTTTCTTACC TATAAAATGG 3000  
TGATAAAAAC CGACAGGTTG TTCAAAGGCC CAGATCAGCT AAAGCATGTA 3050  
TATAAGAGCA CGTTGTAAAC TTGAAAGAGA CAAAGGCACA AATGTGGCTG 3100  
TTGATTAATT TGA CTGCTTC TCGTTGCTCG TCACCTCCAT GCCAGGCACT 3150  
GTGCTTGCTA ATTGCTTTAT GGGGGCATT CTTATTTAT TCCCCAGCCC 3200  
TGGGAAATAG GAGCTGTCAT TATCCTTCTC TTTCTGCACA AGGAAAAATT 3250  
AATGCCCTGA GAATTGTCAT AATTTTCCCA AGGCTGCCCCA GCTGGTGGTG 3300  
TTAAGCCAGA ATTTGACCTC CCAGAGCCAG TTTCCATTAG CTGCCATGCT 3350  
CTGCTGCCTC TAATTCACAG AATGCACTTT CTACCCTGTG TGCCATGGAG 3400  
ACCTCCTATG GAAAAATGAT CAGCCACCTT ACCTTCTACT GGGTACCTGC 3450

TGTGAGTCTG CCTATGCCAG AAGGATTAAG GAGGGGAGGT TACCCAAGAA 3500  
 ACAAAGCCTA CATGCCGCTT ACAGCCCCCG TTGGATGGTT GCTCAGTACA 3550  
 ACAGTCTTGC ATTCAGCAGG TGTTTGTTCA TCACCTACTA TGTGTCAGGC 3600  
 TCTATGCTAG GTACTGGGGA TACAGGAGAG AATCAAGCGT AAAGTCTTTG 3650  
 TTCTCAAGGA ATTTGCATTC TAGAAAGTAG AAGATGTAAT AAATGTACTG 3700  
 TGGGACATGT TAATAAGTGC TATAAAGAAA TATAAAGGGT TTGGGAGCAA 3750  
 AAAGAGGGAG TGGATCTATT TTAGATGAGC CCAGGTAAGA CCTCTCTGAA 3800  
 GAGCTGTCAT GAAGGAGGGA GGGAGCACAT TCCTGGCAGA GAAAACAGCA 3850  
 CGTGCAAAGG CCCCAGAGCT GGAGTGTGTT CCTGAAGAGC AGCCAGGAGG 3900  
 CCAGCATGGC TGGAGAGGCA GGCATAGGCA GGAACCGAG CAGCAGGTCA 3950  
 GAGCAGGCGA GCTGACATTC TGCAGCCTGG ACGGCCATGG CAGGAAGCTT 4000  
 TTAGTTGGAG AGATACAGGA AGCCTCCTAG GGTTCTGAGC AGAAGAGGGG 4050  
 CATGAGCTGA TTCACATTCT GAAGGACCTC TCTAGCTGGC CAGTGCTGAG 4100  
 GAGGTTGGAG AGAGAAAGGG TGAAAGCAGA GAGACCAGTG CAGGGCTGTT 4150  
 AACAGGGTTG CAGGCGAGAG ACTGGGGTGC TGGGCTCCCC TAGACTAGGA 4200  
 CTCCAGTGCC CTCCTCTCCC AAGAGACAAA GGCCATTGCA TTGAAGGAGG 4250  
 TGGGAAATGA TTAGATTCTG AACATATGTA ATTATTTTTC AGTCTTTTTC 4300  
 AAAGATACAA ATATTTACAT AGTTTAAATC ATGTAATATA TACAATTTAA 4350  
 TGTCCTAGTG TTTTACTTAA TAGTGTATCA TGTTTTCCCT GTTGGTATGT 4400  
 AGCCTGGATA AATGCTCTTA ATTATAAAAA ATTCTGTCGA GGAGTGTTCC 4450  
 ATAGTTTATT GTTTTCCTAT TATGAGAATT TAGGCCAAGT GTGGTGGCTC 4500

ATGCCTGTAA TCCCAGCACT TTGCGAGGCC GAGGTGGGCA GATCACTTGA 4550  
 GGTGAGGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAATT ATCTCTACTA 4600  
 AAAATACAAA AAAATAATAA TAATAGCCAG GCGTGGTGGC ACATGCCTGT 4650  
 ATTCCCAGCT GCTTGGGAGG CTGAGGCAGG AGAATGGCTT GAACCTGGGA 4700  
 GGTGGAGGTT GCAGTGAGCC GAGATGGTGC CACTGCATTC CAGCCTGGGC 4750  
 AACAGAGCGA GACTCCATCT CAAAAAAAAG GAGACTTCAT GTGCCCCCAA 4800  
 TTTTTCATA TTGTTATTTG AAAAAATATT TTTATTTGTA AGAGTTTTTC 4850  
 TTTATTTAAA ATGTTCATTA ATAAAGTTGT TGGACGGGAA GCAAAAAAAA 4900  
 AAAGTTGTTT AAGATAAATT CCCAGAAGTG AATTGTAG ATCAAACACT 4950  
 TAAAACTTTT TGTATGGAA GAATTCAAAT ATAAATAAAA AATTGTGAGT 5000  
 AATAAAATG 5009

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 474 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Ser	Asn	Tyr	Arg	Tyr	Ser	Lys	Leu	Asn	Glu	Glu	Glu	Ile	Ser
1				5					10					15
Leu	Glu	Asp	Met	Pro	Ser	Ser	Ala	Asn	Gln	Ile	Leu	Thr	Arg	Gln
				20					25					30
Glu	Gln	Ile	Ile	Gln	Glu	Gln	Asp	Asp	Glu	Leu	Glu	Leu	Val	Gly
				35					40					45
Asn	Ser	Val	Arg	Thr	Leu	Arg	Gly	Met	Ser	Ser	Met	Ile	Gly	Asp
				50					55					60

Glu	Leu	Asp	Gln	Gln	Ser	Thr	Met	Leu	Asp	Asp	Leu	Gly	Gln	Glu	65	70	75
Met	Glu	Tyr	Ser	Glu	Thr	Arg	Leu	Asp	Thr	Ala	Met	Lys	Lys	Met	80	85	90
Ala	Lys	Leu	Thr	His	Leu	Glu	Asp	Gly	Met	Leu	Leu	Ala	Arg	Arg	95	100	105
Ile	Val	Gln	Ser	Met	Gln	Asn	Asp	His	Gly	Ala	Leu	Ser	Ser	Pro	110	115	120
Val	Phe	Pro	Arg	Leu	Cys	Pro	Ser	Gly	Leu	Thr	Thr	Tyr	Val	Pro	125	130	135
Tyr	Ile	Val	Asp	Phe	Ser	Ser	Leu	Thr	Phe	His	Ile	Phe	Ile	Ile	140	145	150
Ile	Ile	Ile	Ile	Ile	Ile	Asp	Phe	Cys	Ser	Gln	Ser	Gln	Ser	Lys	155	160	165
Gly	Arg	Phe	Gly	Asn	Gln	Val	Asp	Gln	Phe	Leu	Gly	Val	Leu	Ala	170	175	180
Phe	Ala	Lys	Ala	Leu	Asp	Arg	Thr	Leu	Val	Leu	Pro	Asn	Phe	Ile	185	190	195
Glu	Phe	Lys	His	Pro	Glu	Thr	Lys	Met	Ile	Pro	Phe	Glu	Phe	Leu	200	205	210
Phe	Gln	Val	Gly	Thr	Val	Ala	Lys	Tyr	Thr	Arg	Val	Val	Thr	Met	215	220	225
Gln	Glu	Phe	Thr	Lys	Lys	Ile	Met	Pro	Thr	His	Phe	Val	Gly	Thr	230	235	240
Pro	Arg	Gln	Ala	Ile	Tyr	Asp	Lys	Ser	Ala	Glu	Pro	Gly	Cys	His	245	250	255
Ser	Lys	Glu	Gly	Asn	Pro	Phe	Gly	Pro	Tyr	Trp	Asp	Gln	Ile	Asp	260	265	270

Val	Ser	Phe	Val	Gly	Asp	Glu	Tyr	Phe	Gly	Asp	Ile	Pro	Gly	Gly			
				275					280					285			
Phe	Asp	Leu	Asn	Gln	Met	Gly	Ser	Arg	Lys	Lys	Trp	Leu	Glu	Lys			
				290					295					300			
Phe	Pro	Ser	Glu	Glu	Tyr	Pro	Val	Leu	Ala	Phe	Ser	Ser	Ala	Pro			
				305					310					315			
Ala	Pro	Phe	Pro	Ser	Lys	Gly	Lys	Val	Trp	Ser	Ile	Gln	Lys	Tyr			
				320					325					330			
Leu	Arg	Trp	Ser	Ser	Arg	Ile	Thr	Glu	Gln	Ala	Lys	Lys	Phe	Ile			
				335					340					345			
Ser	Ala	Asn	Leu	Ala	Lys	Pro	Phe	Val	Ala	Val	His	Leu	Arg	Asn			
				350					355					360			
Asp	Ala	Asp	Trp	Val	Arg	Val	Cys	Glu	His	Ile	Asp	Thr	Thr	Thr			
				365					370					375			
Asn	Arg	Pro	Leu	Phe	Ala	Ser	Glu	Gln	Cys	Leu	Gly	Glu	Gly	His			
				380					385					390			
His	Leu	Gly	Thr	Leu	Thr	Lys	Glu	Ile	Cys	Ser	Pro	Ser	Lys	Gln			
				395					400					405			
Gln	Ile	Leu	Glu	Gln	Ile	Glu	Ala	His	Arg	Gln	Glu	Pro	Asp	Asp			
				410					415					420			
Met	Tyr	Thr	Ser	Leu	Ala	Ile	Met	Gly	Arg	Ala	Asp	Leu	Phe	Val			
				425					430					435			
Gly	Asn	Cys	Val	Ser	Thr	Phe	Ser	His	Ile	Val	Lys	Arg	Glu	Arg			
				440					445					450			
Asp	His	Ala	Gly	Gln	Ser	Pro	Arg	Pro	Ser	Ala	Phe	Phe	Gly	Ile			
				455					460					465			
Arg	Ala	Val	Lys	Arg	His	Ile	Asp	Leu									
				470				474									

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met	Pro	Ala	Gly	Ser	Trp	Asp	Pro	Ala	Gly	Tyr	Leu	Leu	Tyr	Cys
1				5					10					15
Pro	Cys	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe	Leu	Gly
				20					25					30
Ser	Leu	Ala	Phe	Ala	Lys	Leu	Leu	Asn	Arg	Thr	Leu	Ala	Val	Pro
				35					40					45
Pro	Trp	Ile	Glu	Tyr	Gln	His	His	Lys	Pro	Pro	Phe	Thr	Asn	Leu
				50					55					60
His														
61														

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTTCTTGGGC TCTCTGGCAT TTGCAAAGCT GCTAAACCGT 40

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTCGACGATT TGGCATGGAA CCGACAGGGA GGAACCTAAC 40

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTTCTTGGGC TCTCTGGCAT TTGCAAAGCT GCTAAACCGT 40

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCCCTGGGGA GTTCCTCCCT CTGCGAGGTA 30

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Arg	Ser	His	His	His	His	His	His	Met	Pro	Ala	Gly	Ser	Trp	Asp
1				5					10					15

Pro Ala Gly Tyr Leu Leu Tyr Xaa Pro Xaa Met Gly Arg  
20 25 28

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 52 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Thr	Val	Asp	Gly	Asp	Gln	Cys	Glu	Ser	Asn	Pro	Cys	Leu	Asn	Gly
1				5					10					15
Gly	Ser	Cys	Lys	Asp	Asp	Ile	Asn	Ser	Tyr	Glu	Cys	Trp	Cys	Pro
				20				25						30
Phe	Gly	Phe	Glu	Gly	Lys	Asn	Cys	Glu	Leu	Asp	Val	Thr	His	His
				35				40						45
His	His	His	His	Gly	Ser	Ala								
				50		52								

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1100 base pairs
  - (B) TYPE: Nucleic Acid
  - (C) STRANDEDNESS: Single
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATGCCCCGCGG GCTCCTGGGA CCCGGCCGGT TACCTGCTCT ACTGCCCCTG 50  
CATGGGGCGC TTTGGGAACC AGGCCGATCA CTTCTTGGGC TCTCTGGCAT 100  
TTGCAAAGCT GCTAAACCGT ACCTTGCTG TCCCTCCTTG GATTGAGTAC 150  
CAGCATCACA AGCCTCCTTT CACCAACCTC CATGTGTCCT ACCAGAAGTA 200

CTTCAAGCTG GAGCCCCTCC AGGCTTACCA TCGGGTCATC AGCTTGGAGG 250  
 ATTTTCATGGA GAAGCTGGCA CCCACCCACT GGCCCCCTGA GAAGCGGGTG 300  
 GCATACTGCT TTGAGGTGGC AGCCCAGCGA AGCCCAGATA AGAAGACGTG 350  
 CCCCATGAAG GAAGGAAACC CCTTTGGCCC ATTCTGGGAT CAGTTTCATG 400  
 TGAGTTTCAA CAAGTCGGAG CTTTTTACAG GCATTTCCTT CAGTGCTTCC 450  
 TACAGAGAAC AATGGAGCCA GAGATTTTCT CCAAAGGAAC ATCCGGTGCT 500  
 TGCCCTGCCA GGAGCCCCAG CCCAGTTCCC CGTCCTAGAA GAACACAGGC 550  
 CACTACAGAA GTACATGGTA TGGTCAGACG AAATGGTGAA GACGGGAGAG 600  
 GCCCAGATTC ATGCCACCT TGTCCGGCCC TATGTGGGCA TTCATCTGCG 650  
 CATTGGCTCT GACTGGAAGA ACGCCTGTGC CATGCTGAAG GACGGGACTG 700  
 CAGGCTCGCA CTTTCATGGCC TCTCCGCACT GTGTGGGCTA CAGCCGCAGC 750  
 ACAGCGGCCC CCCTCACGAT GACTATGTGC CTGCCTGACC TGAAGGAGAT 800  
 CCAGAGGGCT GTGAAGCTCT GGGTGAGGTC GCTGGATGCC CAGTCGGTCT 850  
 ACGTTGCTAC TGATTCCGAG AGTTATGTGC CTGAGCTCCA ACAGCTCTTC 900  
 AAAGGGAAGG TGAAGGTGGT GAGCCTGAAG CCTGAGGTGG CCCAGGTCGA 950  
 CCTGTACATC CTCGGCCAAG CCGACCACTT TATTGGCAAC TGTGTCTCCT 1000  
 CCTTCACTGC CTTTGTGAAG CGGGAGCGGG ACCTCCAGGG GAGGCCGTCT 1050  
 TCTTTCTTCG GCATGGACAG GCCCCCTAAG CTGCGGGACG AGTTCTGATT 1100

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 343 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Asn	Gln	Ala	Asp	His	Phe	Leu	Gly	Ser	Leu	Ala	Phe	Ala	Lys	Leu	
1				5					10					15	
Leu	Asn	Arg	Thr	Leu	Ala	Val	Pro	Pro	Trp	Ile	Glu	Tyr	Gln	His	
				20					25					30	
His	Lys	Pro	Pro	Phe	Thr	Asn	Leu	His	Val	Ser	Tyr	Gln	Lys	Tyr	
				35					40					45	
Phe	Lys	Leu	Glu	Pro	Leu	Gln	Ala	Tyr	His	Arg	Val	Ile	Ser	Leu	
				50					55					60	
Glu	Asp	Phe	Met	Glu	Lys	Leu	Ala	Pro	Thr	His	Trp	Pro	Pro	Glu	
				65					70					75	
Lys	Arg	Val	Ala	Tyr	Cys	Phe	Glu	Val	Ala	Ala	Gln	Arg	Ser	Pro	
				80					85					90	
Asp	Lys	Lys	Thr	Cys	Pro	Met	Lys	Glu	Gly	Asn	Pro	Phe	Gly	Pro	
				95					100					105	
Phe	Trp	Asp	Gln	Phe	His	Val	Ser	Phe	Asn	Lys	Ser	Glu	Leu	Phe	
				110					115					120	
Thr	Gly	Ile	Ser	Phe	Ser	Ala	Ser	Tyr	Arg	Glu	Gln	Trp	Ser	Gln	
				125					130					135	
Arg	Phe	Ser	Pro	Lys	Glu	His	Pro	Val	Leu	Ala	Leu	Pro	Gly	Ala	
				140					145					150	
Pro	Ala	Gln	Phe	Pro	Val	Leu	Glu	Glu	His	Arg	Pro	Leu	Gln	Lys	
				155					160					165	
Tyr	Met	Val	Trp	Ser	Asp	Glu	Met	Val	Lys	Thr	Gly	Glu	Ala	Gln	
				170					175					180	
Ile	His	Ala	His	Leu	Val	Arg	Pro	Tyr	Val	Gly	Ile	His	Leu	Arg	
				185					190					195	
Ile	Gly	Ser	Asp	Trp	Lys	Asn	Ala	Cys	Ala	Met	Leu	Lys	Asp	Gly	
				200					205					210	

Thr	Ala	Gly	Ser	His	Phe	Met	Ala	Ser	Pro	Gln	Cys	Val	Gly	Tyr	215	220	225
Ser	Arg	Ser	Thr	Ala	Ala	Pro	Leu	Thr	Met	Thr	Met	Cys	Leu	Pro	230	235	240
Asp	Leu	Lys	Glu	Ile	Gln	Arg	Ala	Val	Lys	Leu	Trp	Val	Arg	Ser	245	250	255
Leu	Asp	Ala	Gln	Ser	Val	Tyr	Val	Ala	Thr	Asp	Ser	Glu	Ser	Tyr	260	265	270
Val	Pro	Glu	Leu	Gln	Gln	Leu	Phe	Lys	Gly	Lys	Val	Lys	Val	Val	275	280	285
Ser	Leu	Lys	Pro	Glu	Val	Ala	Gln	Val	Asp	Leu	Tyr	Ile	Leu	Gly	290	295	300
Gln	Ala	Asp	His	Phe	Ile	Gly	Asn	Cys	Val	Ser	Ser	Phe	Thr	Ala	305	310	315
Phe	Val	Lys	Arg	Glu	Arg	Asp	Leu	Gln	Gly	Arg	Pro	Ser	Ser	Phe	320	325	330
Phe	Gly	Met	Asp	Arg	Pro	Pro	Lys	Leu	Arg	Asp	Glu	Phe			335	340	343